DOCUMENT RESUME

ED 064 348 TM 001 515

AUTHOR Harris, Chester W.
TITLE Interbattery Factor Analysis and Canonical Correlation Analysis as Tools for Relating Concept Attainment Measures and Cognitive Abilities Measures.
PUB DATE Apr 72
NOTE 13p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cognitive Ability; *Concept Formation; *Correlation; *Factor Analysis; *Statistical Analysis; Tests

ABSTRACT

The use of two methods of analysis in studying the relations between two sets of measures--the cognitive abilities tests and the concept attainment measures--was studied. The two methods of analysis are the Canonical Variate Analysis and the Interbattery Factor Analysis. The proposed application of each method is described. A summary of the computations that are envisioned is given. (DB)

ERIC

ED 064348

INTERBATTERY FACTOR ANALYSIS AND CANONICAL CORRELATION ANALYSIS AS TOOLS FOR RELATING CONCEPT ATTAINMENT MEASURES AND COGNITIVE ABILITIES MEASURES

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Chester W. Harris
University of California, Santa Barbara

AERA
Chicago, Illinois
April 4, 1972

TM 001 515

A final problem is that of studying the relations between two sets of measures: the cognitive abilities tests and the concept attainment measures. The major purpose of such a study is to attempt to identify those cognitive abilities that are related to concept attainment in the four subject matter areas. If successful, such an identification should provide a better understanding of which cognitive abilities may be key abilities in the learning of concepts, and which cognitive abilities are differentially related (if they are) to concept attainment in the four subject matter areas. Such information should be of value in designing experimental studies of instruction in the four subject matter areas. I am reporting today our plans for the analysis. The actual results are not yet available and will be reported subsequently.

In studying the relations between two sets of variables, one may adopt three somewhat different attitudes. Let the matrix of intercorrelations of one set of variables be labeled R_{xx}, and of the other, R_{yy}. Also let R_{xy} designate the cross correlations between the two sets of variables. These may be arranged as a supermatrix:

$$\begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}$$

in which R_{xx} and R_{yy} are symmetric, but R_{xy} and its transpose, R_{yx}, are (generally) rectangular. As one example, our 1971 data provides intercorrelations of the 31 cognitive abilities tests (R_{xx}), intercorrelations of the 48 task scores (R_{yy}), and the cross correlations of 31 cognitive abilities tests with 48 task scores (R_{xy}).

Given such a supermatrix, one might ignore the distinction between the two sets of variables and simply factor the entire matrix. One would then hope that factors of such a matrix could be transformed ("rotated") to reveal distinctive patterns of linkages between the two sets of variables. We felt that this attitude was likely to be less productive of informative results than others we might adopt.

A second attitude would capitalize on the distinction between the two batteries, but would regard one of the two batteries as fundamental in some sense, and would first analyze the "fundamental" battery into factors (or possibly components). The question of the linkage of the variables of the second battery with the factors of the fundamental battery could then be answered by computing the correlations of these variables with the factors of the fundamental battery. An example of this approach is given by Bunderson (1965) who regarded the cognitive abilities as measured by the tests in the ETS Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963) as making up the variables of the fundamental battery; he studied measures of concept attainment as the second battery. For our problem, it would be conventional to regard "aptitude" as fundamental and "achievement" as (at least in part) a manifestation of "aptitude," and thus to regard the cognitive abilities tests as making up the fundamental battery for such an analysis.

We preferred, however, to adopt a third attitude in which we regard the two batteries as of equal importance _a priori_ and then employ methods of analysis which attend to the distinction between the two batteries, but do not necessarily regard one of the two as fundamental. Our choice was to use two related procedures--Canonical Variate Analysis (Hotelling, 1935) and Interbattery Factor Analysis (Tucker, 1958).

CANONICAL VARIATE ANALYSIS

Hotelling (1935) showed that for any supermatrix

$$\begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}$$

it is possible to find a linear composite of the x variables and a linear composite of the y variables such that these two composites have a largest possible correlation. (Multiple correlation is a special case, in which one of the two batteries consists of only one variable.) Further, a second pair of composites can then be found that are uncorrelated with the first pair of composites but have a next largest correlation, and so forth. The number of pairs cannot exceed the number of variables in the smaller battery.

These pairs of composites that are maximally correlated pairwise, but uncorrelated otherwise, may be called canonical variates. Of major interest are: the pairwise correlations, which are called canonical correlations; the weights on the variables or weight matrices that define these composites; and the correlations of the variables with the canonical variates.

In general, R_{xx} and R_{yy} are nonsingular matrices when they are developed for linearly independent sets of variables. This will be true for our data. (Linear independence would be violated, as one example, if two part scores from a given test and the sum of these two part scores were all included as variables in a set.) More than one calcualtion routine can be used; we will use the following.

Let $R_{xx} = C_x C_x'$, where C_x is square and nonsingular. This is a resolution of R_{xx} into components and may be done in many ways; a particular solution would be the complete set of principal components. Similarly, let $R_{yy} = C_y C_y'$. It then is possible to compute C_x^{-1} (the inverse of C_x) and C_y^{-1}, and to use these in computing the weight matrices. What is called a determinantal equation is employed in computing the canonical correlations; a pair of these equations exist. They are:

$$\left| R_{yx} R_{xx}^{-1} R_{xy} - \lambda R_{yy} \right| = 0$$

$$\left| R_{xy} R_{yy}^{-1} R_{yx} - \lambda R_{xx} \right| = 0$$

A matrix product like $R_{yx} R_{xx}^{-1} R_{xy}$ is of interest in itself, since it represents the covariances (not correlations) of the linearly predicted values of the y variables when the x variables are used to make the predictions. The variances of these linearly predicted y values are squared multiple correlations; these appear in the principal diagonal of $R_{yx} R_{xx}^{-1} R_{xy}$. A comparison of $R_{yx} R_{xx}^{-1} R_{xy}$ with R_{yy} gives some notion of how effectively the x variables estimate the y variables. A similar analysis can be made for estimating the x variables from the y variables.

It is convenient to solve the determinantal equations by making the linear transformations:

$$\left| C_y^{-1} R_{yx} R_{xx}^{-1} R_{xy} (C_y')^{-1} - \lambda I \right| = 0$$

$$\left| C_x^{-1} R_{xy} R_{yy}^{-1} R_{yx} (C_x')^{-1} - \lambda I \right| = 0$$

The nonzero eigenvalues, λ, of either of these equations are the squared canonical correlations. Let these be the diagonal elements in a diagonal matrix, D^2. Associated with these nonzero eigenvalues is a set of orthonormal eigenvectors, W_y and W_x for the two equations, respectively. Then the weight matrices defining the canonical variates are $(C_y')^{-1}W_y$, giving the weights to be applied to the y variables, and $(C_x')^{-1}W_x$, giving the weights to be applied to the x variables.

It is assumed that the data for either the x or the y variables are in deviation form; in addition it is assumed that the data are scaled so that the intercorrelations are given by the sums of the paired products of these deviated and scaled scores. Let X designate such a set of data, with rows corresponding to individuals in the sample. Then when $X'X = R_{xx}$, it can be shown that these canonical variates are uncorrelated and are scaled as unit length vectors. Then the correlations of the two sets of original variables with the two sets of canonical variates are given by C_xW_x, $R_{yx}(C_x')^{-1}W_x$, C_yW_y, and $R_{xy}(C_y')^{-1}W_y$. This first matrix consists of correlations of the x variables with the canonical variates that are composites of the x variables. This second matrix consists of correlations of the y variables with the canonical variates that are composites of the x variables. Similarly, the third matrix consists of correlations of the y variables with the canonical variates that are composites of the y variables, and the fourth one consists of correlations of the x variables with the canonical variates that are composites of the y variables. For each of these matrices the rows correspond to the original variables and the columns to the canonical variates. One may then verify that $W_x'C_x^{-1}R_{xy}(C_y')^{-1}W_y$ is a diagonal matrix D giving the nonzero canonical correlations between the two sets of variables. The matrices $C_xW_xD^{\frac{1}{2}}$ and $C_yW_yD^{\frac{1}{2}}$ are rescalings of the columns of the correlations of the original variables with related canonical variates.

In addition $(C_xW_xD^{\frac{1}{2}})\ (C_yW_yD^{\frac{1}{2}})' = C_xW_xDW_y'C_y'$ which reproduces R_{xy}, and so $C_xW_xD^{\frac{1}{2}}$ and $C_yW_yD^{\frac{1}{2}}$ are analogs of what are described below as Interbattery Factors.

These analyses give the nonzero canonical correlations between the two sets of variables, the covariances of the estimates of each set of variables based on the other set of variables, the correlations of each set of variables with the canonical variates, and analogs of Interbattery Factors. In addition, there is available a test which indicates the number of statistically significant canonical correlations, and thus gives guidance in determining how many canonical variates should be interpreted as nonchance variables.

INTERBATTERY FACTOR ANALYSIS

As originally proposed by Tucker (1958), Interbattery Factor Analysis was devised to provide information relevant to the stability of factors over different selections of tests. Tucker says, "Two batteries of tests, postulated to depend on the same common factors, but not parallel tests, are given to one sample of individuals. Factors are determined from the correlation of the tests in one battery with the tests in the other battery. These factors are only those that are common to the two batteries." (Tucker, 1958, p. 111.) We propose to use Tucker's procedures in a slightly different situation. Our two batteries are not specifically designed to depend upon the same common factors; instead, the question of interest in this study is: Are there factors common to the battery of cognitive abilities tests and the battery of concept attainment measures, and if so, what is their number and nature? Tucker's procedure appears to be especially suitable for answering such a question.

Interbattery Factor Analysis focuses upon the matrix of cross-correlations of the two batteries. This is the matrix R_{xy} (or R_{yx}, its transpose). The object is to derive two matrices, A_x and A_y, such that the product A_xA_y' either reproduces R_{xy} or approximates R_{xy} to some degree of satisfaction. The matrix A_x then gives correlations of the variables of the x battery with the Interbattery Factors, and the matrix A_y gives correlations of the variables of the y battery with these same factors. Having solved for these matrices, one may "rotate" or transform them if he wishes: we will do this in two slightly different ways.

To determine A_x and A_y one may decompose the rectangular matrix R_{xy} in the manner described by Tucker. The number of factors common to the two batteries may range from none (which is the correct number when R_{xy} is null) to an absolute maximum equal to the number of variables in the smaller set of variables. Although for actual data this absolute maximum number usually is required to secure a complete reproduction of R_{xy} from the product A_xA_y', at least some of these factors are likely to be judged to be trivial. One or more guides to the number of Interbattery Factors to extract are needed. We are inclined to use the result of the test for the number of significant canonical correlations as a guide to the minimum number of Interbattery Factors rather than the test described by Tucker. We also will examine the eigenvalues of the summetric product $R_{xy}R_{yx}$, and use the size of the successive eigenvalues as a second guide.

Given A_x and A_y such that A_xA_y' approximates R_{xy} to a satisfactory degree, it may be desirable to "rotate" or transform these two factor matrices to achieve a simpler or clearer interpretation of the factors common to the two batteries. The transformations we propose are subject to the restriction that the same transformation be applied to both matrices, A_x and A_y. This differs from Tucker's practice. One can use the normal varimax procedure in two ways: First, derive the normal varimax transformation based on A_x and then apply this transformation to both A_x and A_y. Second, derive the normal varimax transformation based on the supermatrix

$$\begin{bmatrix} A_x \\ \\ A_y \end{bmatrix}$$

and apply this transformation to the supermatrix. The first method tends to focus on the structure of the Interbattery Factors for the x variables, and as such it tends to regard the x variables as making up the more important battery. The second method is more conventional. Oblique transformations present some interesting theoretical problems which we are not yet ready to discuss.

We list here a summary of the computations that are envisioned.

ERIC

SUMMARY OF COMPUTATIONS

$\begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}$	The supermatrix of correlations for two sets of variables
$\lvert R_{yx}R_{xx}^{-1}R_{xy} - \lambda R_{yy} \rvert = 0$ $\lvert R_{xy}R_{yy}^{-1}R_{yx} - \lambda R_{xx} \rvert = 0$	Determinantal equations
$\lvert C_y^{-1}R_{yx}R_{xx}^{-1}R_{xy}(C_y')^{-1} - \lambda I \rvert = 0$ $\lvert C_x^{-1}R_{xy}R_{yy}^{-1}R_{yx}(C_x')^{-1} - \lambda I \rvert = 0$	Transformed determinantal equations
$R_{xy}R_{yy}^{-1}R_{yx}$	Covariances of the estimated x variables when the estimates are based on the set of y variables
$R_{yx}R_{xx}^{-1}R_{xy}$	Covariances of the estimated y variables, when the estimates are based on the set of x variables
$(C_x')^{-1}W_x$	Weights for deriving the x-canonical variates from the x variables
$(C_y')^{-1}W_y$	Weights for deriving the y-canonical variates from the y variables
C_xW_x	Correlations of x variables with the x-canonical variates
C_yW_y	Correlations of y variables with the y-canonical variates
$R_{xy}(C_y')^{-1}W_y$	Correlations of x variables with the y-canonical variates
$R_{yx}(C_x')^{-1}W_x$	Correlations of y variables with the x-canonical variates
$W_y'C_y^{-1}R_{yx}(C_x')^{-1}W_x$ or $W_x'C_x^{-1}R_{xy}(C_y')^{-1} W_y$	Correlations of x-canonical variates with the y-canonical variates, symbolized by the diagonal matrix D (the canonical correlations)

$C_xW_xD^{\frac{1}{2}}$	An analog of Interbattery Factors for x variables
$C_yW_yD^{\frac{1}{2}}$	An analog of Interbattery Factors for y variables
A_x	Correlations of x variables with Interbattery Factors
A_y	Correlations of y variables with Interbattery Factors
A_xA_x'	Covariances of estimated x variables when the estimates are based on the Interbattery Factors
A_yA_y'	Covariances of estimated y variables when the estimates are based on the Interbattery Factors
A_xA_y'	Cross-covariances of estimated x and estimated y variables when estimates are based on Interbattery Factors (an approximation to R_{xy})
$P_x'R_{xx}P_x$	Covariances of Interbattery x-composites, which are analogs of x-canonical variates (Here $A_x = P_xL$, with L a diagonal matrix and P_x a set of orthonormal columns)
$P_y'R_{yy}P_y$	Covariances of Interbattery y-composites ($A_y = P_yL$, with L diagonal as before and P_y a set of orthonormal columns)
$R_{xx}P_x$	Covariances of x-variables with the Interbattery x-composites (These are analogous to $C_xW_x = R_{xx}(C_x')^{-1}W_x$)
$R_{yy}P_y$	Covariances of y-variables with the Interbattery y-composites (These are analogous to $C_yW_y = R_{yy}(C_y')^{-1}W_y$)
$R_{xy} - A_xA_y'$	Residual cross-covariances after Interbattery Factors have been extracted
$R_{xx} - A_xA_x'$	Residual covariances of x variables after Interbattery Factors have been extracted

$R_{yy} - A_yA_y'$	Residual covariances of y variables after Interbattery Factors have been extracted
$A_yA_x'R_{xx}^{-1}A_xA_y'$	Covariances of the estimated y variables, when the estimates are based on the set of x variables but $A_yA_x'R_{xx}^{-1}$ is used as the set of regression weights in place of $R_{yx}R_{xx}^{-1}$ (This takes A_yA_x' as a satisfactory approximation to R_{yx})
A_xT_1 A_yT_1	Rotated Interbattery Factors, using the first method for developing the transformation
A_xT_2 A_yT_2	Rotated Interbattery Factors, using the second method

REFERENCES

Bunderson, C. V. Transfer of mental abilities at different stages of practice in the solution of concept problems. Unpublished doctoral dissertation, Princeton University, 1965.

French, John W., Ekstrom, Ruth B., and Price, Leighton A. *Kit of reference tests for cognitive factors*. Princeton: Educational Testing Service, 1963.

Hotelling, Harold. The most predictable criterion. *Journal of Educational Psychology*, 1935, *26*, 139-142.

Tucker, Ledyard R. An inter-battery method of factor analysis. *Psychometrika*, 1958, *23*, 111-136.